Recent change in the productivity and schematicity of the way-construction: a distributional semantic analysis

Perek, Florent

DOI:
10.1515/cllt-2016-0014

Document Version
Peer reviewed version

Citation for published version (Harvard):
Perek, F 2016, 'Recent change in the productivity and schematicity of the way-construction: a distributional semantic analysis', *Corpus Linguistics and Linguistic Theory*. https://doi.org/10.1515/cllt-2016-0014

Recent change in the productivity and schematicity of the *way*-construction: a distributional semantic analysis

Florent Perek
University of Birmingham

This paper presents a corpus-based study of recent change in the English *way*-construction, drawing on data from the 1830s to the 2000s. Semantic change in the distribution of the construction is characterized by means of a distributional semantic model, which captures semantic similarity between verbs through their co-occurrence frequency with other words in the corpus. By plotting and comparing the semantic domain of the three senses of the construction at different points in time, it is found that they all have gained in semantic diversity. These findings are interpreted in terms of increases in schematicity, either of the verb slot or the motion component contributed by the construction.

Keywords: productivity, *way*-construction, recent change, diachronic construction grammar, schematicity

1. Introduction

This paper reports on a corpus-based study conducted within the framework of diachronic construction grammar (Bergs & Diewald 2008; Fried 2009, 2010, 2013; Hilpert 2013; Noël 2007; Traugott 2003, 2008; Traugott & Trousdale 2013; Trousdale 2010, 2012, *inter alia*).[1] In this relatively recent field, language change is considered with reference to the idea that grammar should be described as a structured inventory of form-meaning pairs, aka constructions (Fillmore et al. 1988; Fried & Östman 2004; Goldberg 1995, 2006; Kay & Fillmore 1999). Constructions are linked in a network, and may capture grammatical patterns at any level of complexity and abstraction, including specific exemplars. Examples of constructions include word formation rules (e.g., *un-*, *-ness*), partially filled idioms and other multi-word expressions (e.g., *kick the bucket*, *pull X's leg*, *the X-er, the Y-er*), as well as fully abstract phrasal constructions (e.g., the ditransitive construction: NP V NP NP, *He gave her a book*). In diachronic construction grammar, diachronic variation is characterized in terms of changes in particular constructions (constructional changes, cf. Fried 2013; Hilpert 2013; Traugott & Trousdale 2013), or change in the network of constructions, including the creation of new nodes (constructionalization, cf. Fried 2013; Noël 2007; Traugott & Trousdale 2013). Both theoretically and methodically, diachronic construction grammar is usage-based: it subscribes to the notion that important aspects of grammatical representations are shaped by natural language use (Bybee 2006, 2010), and consequently, language change is analyzed by looking at usage data. In particular, data from earlier periods can be relied on to define two important properties of constructions identified by Traugott & Trousdale (2013): schematicity and productivity. Schematicity refers to the level of detail in the form or the function of constructions, and correspondingly the restrictions that are placed on their

[1] I would like to thank two anonymous reviewers for their useful comments on an earlier version of this paper. I am also indebted to Amanda Patten for her help with the inter-annotator agreement study and her feedback on the article.

instances. Productivity refers to the range of lexical items that may fill the slots of constructions.

Against this backdrop, the present study is concerned with recent change in the *way*-construction (e.g., *They hacked their way through the jungle*), with a particular focus on the productivity of its verb slot. Previous work on the construction and arguments that it qualifies as an independent form-meaning pair are summarized in Section 2. Section 3 discusses the data collected for the present study, drawn from the Corpus of Historical American English (Davies 2012), and reports that the distribution of verbs in all three senses of the construction has increased during the period 1830-2009. Section 4 focuses on describing these developments in more qualitative terms by characterizing the meaning of the verbs in the distribution. A distributional semantic method is used in order to plot graphical representations of the productivity of the construction at different points in time, in which semantic groups of verbs are identified. It is shown that the semantic domain of the construction has significantly expanded over the past 180 years. Implications of these findings for the description of change in the grammatical representation of the construction over time are then discussed.

2. The *way*-construction

The construction considered in the present study is exemplified by the following sentences:[2]

(1) He made his way to the workbench.

(2) [F]armers […] were beginning to hack their way through primeval forests.

(3) [W]e talked our way into the VIP area.

(4) [H]e could watch trainees grunting their way under barbed wire.

Formally, this construction consists of a verb and two post-verbal elements: a noun phrase containing the noun *way* and a possessive determiner co-referential with the subject argument, and an adverbial describing some kind of path (usually a prepositional phrase). Semantically, what all instances of this construction share is that they entail motion of the subject referent along the path described by the adverbial.

In one of the first accounts of the *way*-construction, Jackendoff (1990) (who also seems to have coined the term) already notes many of its peculiar syntactic and semantic properties, which set such sentences apart from superficially similar structures containing a direct object NP and a path phrase (such as *throw the ball over the fence*). Jackendoff suggests that this pattern could be "thought of as a kind of 'constructional idiom', a specialized syntactic form with an idiomatic meaning, marked by the noun *way*" (p. 221). In line with this earlier proposal, Goldberg (1995) analyzes this pattern as a direct, conventional pairing of form with meaning, i.e., a construction in the construction grammar sense (Fillmore et al. 1988; Fried & Östman 2004; Goldberg 1995, 2006; Kay & Fillmore 1999).

[2] Unless indicated otherwise, all examples in this section are from the Corpus of Historical American English (COHA; Davies 2012).

At least two properties warrant the description of this pattern as a construction. First, in many cases, the post-verbal constituents cannot plausibly be licensed by the verb. The construction is attested with verbs that do not normally take a direct object (for instance *talk* in example [3]), and even some of the verbs that do cannot occur with both NP and PP complements, as shown by the following examples taken from Jackendoff (1990: 211).

(5) a. *Bill belched (a belch) out of the restaurant.

b. *Harry moaned (a long serious moan) down the road.

c. *Sam joked (many jokes) into the meeting.

Hence, no general rule of grammar can independently license the occurrence of these verbs in the *way*-construction. Second, the motion interpretation in particular cannot be attributed to any of the construction's component parts. As noted by Goldberg (1995: 199), this interpretation holds regardless of whether the verb normally conveys motion or not, and it cannot be attributed to the noun *way* either, since the same sentences differing only in the choice of determiner do not entail motion, as seen in the contrast in acceptability between examples (6a) and (6b) below (taken from Goldberg 1995: 199–200). Rather, it seems that the motion interpretation results from a specific combination of morphosyntactic and lexical features, since it also does not obtain when *way* is replaced with other semantically similar nouns, such as *route* (cf. [7a] vs. [7b], also taken from Goldberg 1995: 199–200).

(6) a. *Frank found his way to New York, but he hasn't gone yet.

b. Frank found a way to New York, but he hasn't gone yet.

(7) a. *Frank dug his way out of prison, but he hasn't gone yet.

b. Frank dug his escape route out of prison, but he hasn't gone yet.

These data suggest that the motion interpretation is conventionally associated with the whole syntactic pattern, and does not plausibly result from pragmatic implicatures, let alone semantic composition. Since the properties of these sentences do not appear to be predicted by their component parts, an independent pairing of form with meaning needs to be posited, which must be learned by speakers for them to understand and produce instances of this construction.

Israel (1996) distinguishes between three uses of the *way*-construction, exemplified by (8) to (10) below, that differ in terms of the role played by the meaning of the verb in the motion scene described by the construction.

(8) They […] bribed their way into city government.

(9) He stumbled his way to the back of the room.

(10) I got out of the car and smoked my way toward the restaurant.

These examples reflect the distinctions that I will follow in this paper. Example (8) shows what I call here the path-creation sense (which Israel calls 'means'), following Traugott & Trousdale (2013). This use of the construction conveys the idea that the subject referent 'creates' in some way a path for itself and moves along this path (cf. Goldberg 1995: 203–209). The notion of path creation can be understood literally as the actual removal of physical obstacles, as in example (2) above with *hack*, or more abstractly as the mere enablement of motion, which can itself be either concrete, as in (3) above with *talk*, or metaphorical, as in (8), where an institution (*city government*) is construed as a container, and becoming a member of this institution as motion into

this container. In any case, the verb refers to the means whereby the 'path' is created. Goldberg (1995: 204) notes that such sentences typically convey that "the subject moves despite some external difficulty", which motivates the fact that a path creation meaning must be expressed in addition to the motion event itself: in Goldberg's words, "there is some reason why a path needs to be created". However, the difficulty reading should merely be seen as a typical implicature, but not necessarily as part and parcel of the semantics of the construction (i.e., an entailment), or as a restriction on its use (i.e., a presupposition); cf. Szczesniak (2013: 165–166). Example (9) is an instance of the manner sense. Like the path-creation sense, it entails motion of the subject referent, but the verb specifies aspects pertaining to the manner in which motion is performed, such as for instance rate (*inch*, *run*, *speed*), terrain (*swim*, *wade*), difficulty (*stagger*, *trudge*), path shape (*thread*, *wend*), etc. Lastly, sentence (10) exemplifies the incidental-action sense, in which the action described by the verb merely occurs concomitantly with motion, but is not directly related to it, let alone causes or enables it as in the path-creation interpretation.[3]

The diachrony of the *way*-construction was famously investigated by Israel (1996), and later re-examined by Traugott & Trousdale (2013: 76–91). Both accounts mostly focus on explaining how the construction came into existence. Hence, they extensively document the earliest stages of its history, but say comparatively little about later developments, especially in the Late Modern English period. Israel (1996) does note that in the nineteenth century, the construction kept attracting new verbs from an increasingly wider semantic range. In particular, "verbs encoding increasingly indirect ways of reaching a goal" (p. 224) are becoming more common in the path-creation interpretation, like for instance predicates that "do not depict any physical exertion but rather mark various social and psychological sorts of activity which enable (literal or metaphorical) motion" (e.g., *smirk*), or "where the overtly coded action only incidentally enables motion, the causal link is even more indirect" (e.g., *write*). Yet, Israel's characterization of these more recent developments remains fairly general, and a precise inventory of verbs or verb classes that joined the distribution of the construction (and when) is lacking.

Besides, neither Israel nor Traugott & Trousdale report frequency information (either in terms of tokens or types), and while their qualitative analysis of citations from the OED for the former, and examples from diachronic corpora for the latter, is certainly useful to identify grammatical and broad distributional properties of the early construction, the data they report can hardly be claimed to exhaustively represent its earlier usage. Corpus-based studies of this kind have been conducted on several constructions in English and other languages (Barðdal 2008; Colleman 2011; Colleman & De Clerck 2011; Noël 2008; Noël & Colleman 2010), but so far not on the *way*-construction. No large-scale corpus-based study of change in the *way*-

[3] The terminology outlined above and the associated categories differ somewhat from other descriptions of the construction found in the literature. In particular, Goldberg (1995) only distinguishes between a 'means' and a 'manner' sense, following a distinction which echoes Jackendoff's (1990) observation that instances of the *way*-construction may fall into two distinct paraphrase relations: some can be paraphrased by "X go Y by V-ing", others by "X go Y while V-ing". In other words, Goldberg's means and manner senses essentially correspond to what I call path-creation and incidental-action respectively. Moreover, she seems to count instances of the construction with manner-of-motion verbs, such as *inch*, *thread*, and *wind*, as instances of the means sense, while I follow Israel (1996) in keeping them as a separate category.

construction (and recent change in particular) has been undertaken to this day, and this paper seeks to mend at least this gap. One possible exception to the latter claim is a study by Mondorf (2011) based on diachronic corpora from the period 1460 to 1900, and present-day English data from the BNC. However, the main focus of Mondorf's study is the competition between the *way*-construction and the *self*-resultative construction (e.g., *She worked herself to the top*), and not the description of the former's usage as such. Because of this particular focus, she considers only verbs that are attested in both constructions (and not just in the *way*-construction), and for reasons of feasibility, her study is limited to ten randomly selected verbs, and thus is not concerned with the total range of verbs occurring in either construction at different points in time.

According to previous studies, the *way*-construction has essentially existed in its modern form from the late 18th century, in that no major qualitative changes are reported to have occurred since then, and the grammatical properties of the constructions have been, by and large, stable. This makes it a prime example to study changes in productivity, i.e., the range of lexical items used in a construction, since no confounding factors related to properties of the construction are present that could have an impact on productivity.

3. Data

The present study uses the offline version of the Corpus of Historical American English (hereafter COHA; Davies 2012). The corpus contains about 385 million words of American English sampled from each decade between 1810 and 2009. Davies (2012) argues that, due to its size and structure, the corpus is especially appropriate for studying the historical development of a wide range of low- and mid-frequency phenomena, for which earlier alternatives (such as the Brown family of corpora) fail to provide robust quantitative data. For this reason, the corpus is well suited to the study of many syntactic constructions, and the *way*-construction in particular.

All instances of a verb followed by a possessive determiner (*my*, *your*, *his*, *her*, *its*, *our*, *their*), the noun *way*, and a preposition, were extracted from the corpus. From these, only instances from the time period 1830-2009 were kept, leaving 20,197 tokens; the 1810s and 1820s were removed from consideration for reasons of data representativeness, since these two decades are markedly smaller than later ones and are less well balanced in genre. In this dataset, instances of the *way*-construction were manually identified. To qualify as such, sentences had to match the semantic and syntactic properties of the construction as defined in Section 2: (i) the possessive had to be co-referential with the sentence subject, (ii) the whole clause had to entail motion of the subject referent (literally or metaphorically), (iii) the preposition had to be the head of a prepositional phrase describing the path of motion, and (iv) the verb had to somehow encode the means whereby motion is enabled, the manner of motion, or an incidental action performed by the subject referent during motion. These criteria ruled out sentences in which *way* is used in an adverbial phrase (e.g., *He came our way*), or is used literally in the sense of 'route' or 'travel' (also sometimes in the abstract sense of 'practice'), which is usually the case with such verbs as *ask*/*inquire*, *know*, *learn*, *see*, *lose*, *miss*, *continue*/*pursue*/*resume*, *start*, *stop*, *light*, *bar*/*block*, among others. The latter uses, exemplified by (11a-c) below, are fully compositional,

and thus are not captured by the *way*-construction; besides, depending on the verb, they do not necessarily entail motion. Similarly, idiomatic expressions such as *have*/*get one's way* are ruled out.

(11) a. [T]he lovers pursued their way across the narrow peninsula.

b. I could barely see my way between the empty graves.

c. Has she lost her way among the narrow, interminable defiles of your crooked old city streets?

Finally, duplicate sentences were removed from the data to the extent that they could be detected, as well as cases whose interpretation was unclear. The manual search returned 17,972 tokens of the *way*-construction in total. This dataset was then semantically annotated by identifying the meaning of the construction in each token, chosen amongst path-creation, manner, or incidental-action, as described in Section 2. To qualify as an instance of the path-creation sense, the verb had to convey the means whereby the motion of the subject referent is enabled or caused, because it literally refers to the creation of an actual path through a physical obstacle, or to the abstract way that some goal is achieved. In other words, the path, or possibility of motion, does not exist before the action described by the verb is performed. If this did not hold, the sentence was classified as an instance of one of the other two senses of the construction. If the verb referred to a certain manner of motion, specifying for instance rate (*inch*, *run*), shape of path (*thread*, *wind*), or shape of motion (*squirm*, *trudge*), the meaning of the token was annotated as the manner sense. If the verb referred to an action unrelated to motion, the token was annotated as an instance of the incidental-action sense. The reliability of the sense annotations was controlled by extracting a random sample of 200 tokens and having them annotated by a second linguist annotator. As part of the instructions for the annotation task, the second annotator was given excerpts of a draft version of this paper, including the sense definitions given in Section 2, all of the examples referred to therein, and the present paragraph describing how sense annotations were performed. There was high (94%) agreement between the annotator and the author, which was confirmed by Cohen's Kappa ($\kappa = 0.835$, $z = 15.2$, $p = 0$).

Figure 1 reports the diachronic variation of four quantitative measures of the three senses of the *way*-construction.[4] The first chart (top left) reports variation in token frequency, i.e., the number of instances of the construction occurring in each decade, normalized per million words. Apart from a recent increase of the path-creation sense from the 1980s onwards, none of the senses shows particularly strong variation in token frequency. The path-creation sense is by far the most common at all times, while the incidental-action sense is clearly still a novelty in the first half of the 19th century, and remains marginal throughout the whole period of interest, despite a steady increase in frequency starting in the early 20th century. This result lines up with Goldberg's (1995: 203) earlier finding that this use of the construction accounts for less than 4% of the tokens of the construction in the contemporary corpora she consulted.

[4] The same trends were found across the four genres in the corpus, showing that the observed changes are not due to genre-specific developments. For reasons of space, the genre analysis is not reported in this paper.

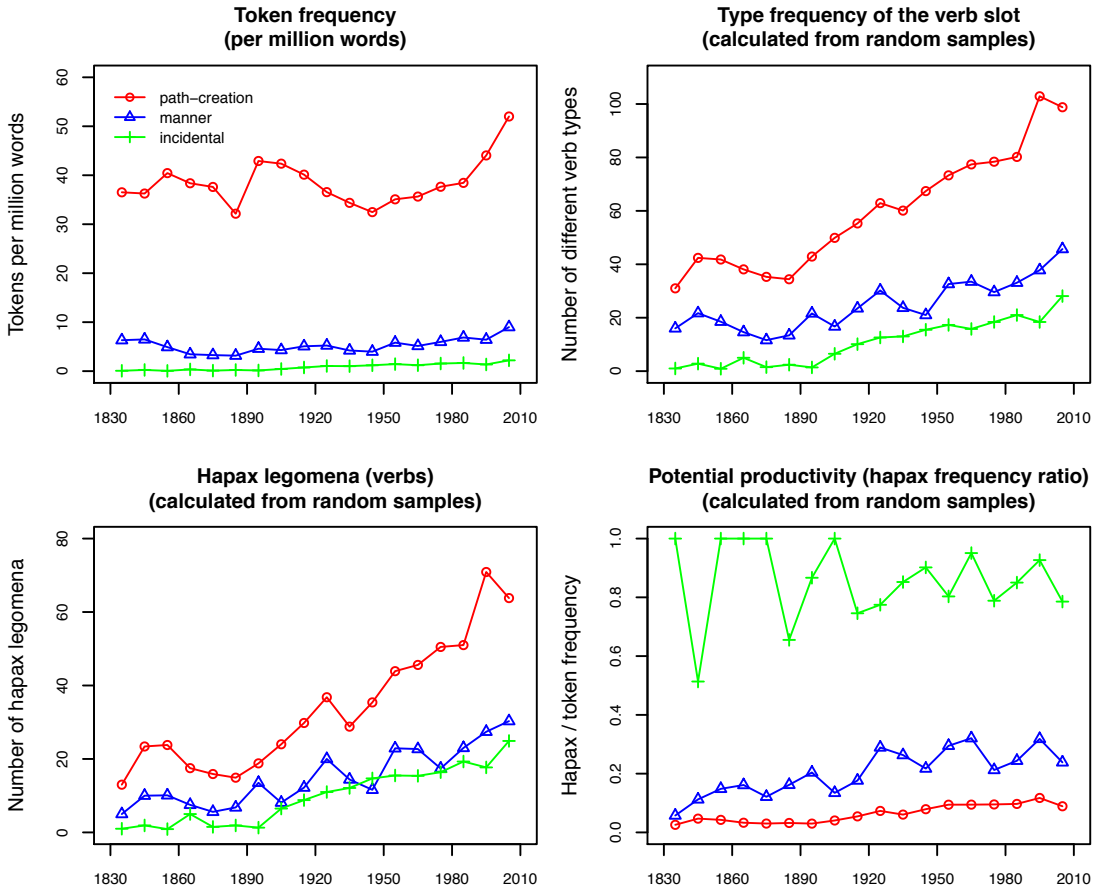

Figure 1: Diachronic variation of four quantitative measures of each sense of the *way*-construction: normalized token frequency, type frequency of the verb slot, number of hapax legomena in the verb slot, and potential productivity of verbs (number of hapax legomena divided by token frequency of the construction). The last three were calculated by averaging over the data from ten random samples of 13.1 million words for each decade.

The next two charts plot the variation in the type frequency of verbs (top right), i.e., the number of different verb stems attested in the construction, and in the number of hapax legomena (bottom left), i.e., verbs that occur only once in the construction. Type frequency is often cited as an indicator of productivity in morphology, where it refers to the number of different words found in a given sample of text that result from the same word formation process. More specifically, type frequency captures what Baayen (2009) terms realized productivity, as opposed to expanding productivity, which reflects the potential of a morphological construction to produce new coinages, and is commonly measured by the number of hapax legomena. The same idea can be extended to slots of a syntactic construction. Baayen & Lieber (1991) suggest another measure of productivity based on hapax legomena, later called potential productivity by Baayen (2009), amounts to the ratio of the number of hapax legomena and the token frequency of the construction; in other words, it corresponds to the proportion of tokens of the construction that occur only once with some type (see also Baayen 1992, 1993 for further discussion). As Baayen & Lieber (1991: 809) put it, this measure captures "the rate at which new types are to be expected to appear when N tokens have been sampled". The variation in potential productivity of the three senses of the *way*-construction is plotted in the fourth chart (bottom left).

The figures reported in Figure 1 were not calculated from the whole corpus, since the amount of data in COHA sampled from each decade between 1830 and 2009 varies quite substantially (between 13.1 and 28.1 million words), which prevents meaningful comparisons of type frequency and hapax legomena across the whole time scale (cf. Gaeta & Ricca, 2006). As is well known since Zipf (1935), the frequency distributions of words in a corpus of natural language are strongly biased towards a few frequent items, and the same was shown to hold for the distribution of syntactic constructions (Ellis et al. 2014; Goldberg et al. 2004; Perek & Lemmens 2010; Römer et al., 2015). This entails that the number of different items in a distribution does not increase linearly with sample size, and there is unfortunately no straightforward way to normalize type frequency like token frequency can be normalized by a simple rule of three. To address this issue, random samples were taken so that every decade matches the smallest size of 13.1 words.[5] For each decade, texts were randomly selected one at a time and their size in number of words tallied, until the total size of the sample reached 13.1 words. Ten randomly sampled sub-corpora were generated for each decade following this method; the corresponding ten sub-samples of the construction were retrieved by restricting the tokens to those occurring in the texts selected in each sample. The counts reported in Figure 1 are average values measured across the ten samples.

As can be seen in Figure 1, all three senses of the *way*-construction have kept attracting more and more new verb types over the past 180 years, especially from the 1890s onwards; the increase is steeper for the path-creation sense than for the other two senses.[6] There is a slight decrease in the last decade (the 2000s), but with the present data it not possible to tell if this indicates the beginning of a trend of decline. At any rate, the general picture across the period of interest is a global increase in realized productivity. A similar observation holds for hapax legomena, showing that the construction has also been open to new coinages. It is worth noting that the incidental-action sense in particular shows remarkably substantial changes, given its humble beginnings. This lines up with Traugott & Trousdale's (2013: 88) suggestion that the incidental-action sense, while recent and infrequent, is comparatively the most productive use of the construction. The hapax/token frequency ratio data further confirms this intuition. The ratio is very unstable in the earlier decades due to the low number of tokens, but it equals one in several decades; in other words, all tokens are

[5] I am grateful to an anonymous reviewer for suggesting this method.

[6] An anonymous reviewer asks whether the increase in types is specific to the various senses of the *way*-construction, or whether it might just correspond to a general increase in the number of verb types in the corpus overall. While this possible confound is not commonly considered in diachronic studies of productivity, this is certainly a valid point that deserves to be addressed. Using the same method as for the calculation of type frequency (i.e., averaging over ten same-size random samples), the total number of different verb types in each decade was extracted from the corpus; to give a rough idea, it varied from 9,318 in the 1830s to 14,606 in the 2000s. The increase in types in each of the four time series (for the three senses and all verbs in the corpus) was quantified by means of least square regression, and by dividing the slope by the initial type frequency in the 1830s, an estimation of the average type growth in each decade relative to the initial number of types was derived. With this method, it was found that type frequency increases by 12.70%, 9.40%, and 147.46% on average each decade for the path-creation, manner, and incidental-action sense respectively, while the average growth in all verb types is only 3.17% per decade. It is thus clear that the increase in type frequency in the *way*-construction cannot be solely accounted for by a general increase in the number of available verb types.

hapax legomena. The ratio stabilizes around 0.8 in later decades, and is much higher than for the other two senses at all times. For the latter, the ratio shows only a slight increase, suggesting that the other senses are steadily productive, but at a relatively lower rate.

All quantitative measures reviewed seem to indicate that the *way*-construction has been very productive in the period under study. However, while quantitative measures of productivity based on type counts are certainly related to the lexical range of a construction, they do not directly capture lexical diversity, i.e., how different the items in the distribution are. As such, they actually provide a rather coarse indication of productivity: with all quantitative aspects kept constant, a construction used with similar items of the same kind can certainly not be considered as productive as a construction attracting highly different items. Many previous studies indicate that semantics plays a central role in syntactic productivity (e.g., Barðdal 2008; Bybee 2010; Bybee & Eddington 2006; Suttle & Goldberg 2011). Therefore, in order to offer a more comprehensive account of the history of the *way*-construction, it is also important to determine (i) what semantic types of verbs joined the distribution of the construction (and when), (ii) whether it became more semantically diverse over time, and (iii) whether there are particular semantic domains favored by the construction. The next section addresses these questions, drawing on a data-driven method based on distributional semantics.

4. Distributional semantic plots of the *way*-construction

4.1. A distributional semantic approach to syntactic productivity

In order to describe the distribution of the *way*-construction, a semantic characterization of the verbs occurring in it is needed. The operationalization of semantics in corpus-based research is not a straightforward issue, since linguistic meaning is not directly observable in the same way that morphosyntactic or phonetic properties are. Most existing studies in the nascent field of diachronic construction grammar rely on the semantic intuitions of the analyst. This paper uses a different approach based on distributional semantics (Erk 2012; Lenci 2008; Turney & Pantel 2010). Drawing on the observation that words occurring in similar contexts tend to have related meanings, distributional semantics holds the view that the meaning of words can be characterized through their distribution in large corpora, as epitomized by Firth's (1957: 11) famous statement "[y]ou shall know a word by the company it keeps". For example, the semantic similarity between the verbs *drink* and *sip* can be captured by the fact that they tend to co-occur with a common set of words, like for instance names for beverages (*tea*, *wine*), typical containers (*cup*, *bottle*), or words related to drinking and dining practices (*bar*, *restaurant*, *table*). In sum, in distributional semantics, the semantic similarity between two words is related to how frequently they are found to share collocates in a vast corpus of naturally occurring texts.

The first step in building a distributional semantic model for a given set of words (for instance, all verbs) is to locate these words in a large corpus and count, for all the tokens found, the frequency of co-occurrence with other words within a set context window (e.g., five words to the left and five words to the right). This results in a co-occurrence matrix, with the set of words under consideration as rows, the collocates as columns, and the co-occurrence frequency in each cell. This matrix already

constitutes a distributional semantic model, in that similarity between rows can be taken to reflect semantic similarity between the corresponding words. However, most modern implementation of distributional models usually go one step further by applying transformations to the co-occurrence matrix, which fall into two kinds: weighting and dimensionality reduction. Weighting follows the same logic as in many collocation studies: to apply a statistical association measure (Evert 2005) to the frequency counts, turning them into measures of how typical and distinctive the collocates are for each target word, i.e., to what extent the co-occurrence frequency diverge from what pure chance would predict, considering the frequency of the word and the overall frequency of its collocates. Dimensionality reduction aims for the matrix to contain fewer columns without losing distributional information; it basically removes redundant information by aggregating frequency counts into more general distributional features by means of linear algebra such as singular value decomposition. In addition to making operations on the matrix computationally more tractable, dimensionality reduction also singles out the most informative aspects of word distributions. Each row of the final matrix is a vector (i.e., an array of numerical values) representing the distributional profile of a given word. Under the assumption that semantic distance between words is a function of distributional differences, similarity between rows, which can be quantified by mathematical measures, approximate semantic similarity. To derive semantic similarity, or its converse, semantic distance, the cosine measure is by far the most frequently used in distributional models of word meaning. Its main advantage is that it normalizes for word frequency, in that two words from a different frequency range will be judged similar if their collocates occur with proportionally similar frequencies, even though the raw frequencies of co-occurrence might differ substantially.

Many implementations of distributional semantic models have been shown to correlate positively with human performance on various tasks, such as synonymy judgments, word association, and semantic priming (Andrews et al. 2008; Lund et al. 1995; Landauer et al. 1998). As argued by Perek (2016), the main benefit of distributional semantics over non-quantitative approaches is that it makes the semantic aspects of the analysis data-driven and automatic, and provides a more objective way of grouping lexical items. Moreover, the informal notion of meaning is turned into a quantified representation that can be used for many purposes, such as the derivation of other quantitative measures, visualization, or statistical testing (see Perek 2016 for examples).

The distributional semantic model created for this study contains distributional information for all verbs found in the COHA.[7] To ensure that the model contains enough distributional data for each verb and obtain a reliable assessment of their distributional similarity, only verbs with an overall frequency of at least 1,000 were kept, which totals 2,532 items. The co-occurrence data was extracted from the offline version of the entire COHA corpus. Lemmatized and part-of-speech-tagged collocates

[7] This is a departure from Perek (2016), in which the semantic model was intended to capture semantic differences between the verbs found in the distribution of a specific construction only. The bigger model used here was found to offer better results, probably due to an effect of weighting and/or dimensionality reduction: when compared to the entire verbal lexicon, distributional information for individual verbs tends to capture more meaningful distinctions and similarities than when it is compared to a much lower number of possibly unrelated items. In the latter case, some features tend to be given too much weight because no other item is found in the distribution that is reasonably similar.

were used as columns of the co-occurrence matrix. Only the 10,000 most frequent collocates (according to their overall frequency in the whole corpus) were kept. Several versions of the distributional model were created by manipulating different aspects of the collocate extraction process or the matrix transformation: the size of the context window (two vs. five words to the left/right), the type of collocates considered (only content words: nouns, lexical verbs, adjectives, and adverbs, vs. all words including functions words: prepositions, conjunction, auxiliary verbs, etc.), the weighting scheme (Positive Pointwise Mutual Information vs. Positive Local Mutual Information), dimensionality reduction (reduction to 300 dimensions by means of Singular Value Decomposition vs. no reduction). The combination of these factors led to 16 different models that were compared according to how well the distributional distance that they capture (as measured by the cosine distance) correlate with judgments of semantic distance provided by human informants. Correlation tests were conducted for 222 pairs of verbs taken from the SimLex-999 norming dataset (Hill et al. 2014). The model performing best for all three tests is the one with content words taken from a two-word context window, PPMI weighting, and dimensionality reduction (Pearson's $r = 0.2893$, $p < 0.001$; Kendall's $\tau = 0.1975$, $p < 0.001$; Spearman's $\rho = 0.2773$, $p < 0.001$).[8]

Pairwise semantic similarity scores between all verbs occurring in the *way*-construction at any point in time were calculated from the distributional semantic model, using the cosine measure. Perek (2016) uses the semantic similarity data to produce visual representations of the semantic domain of the *hell*-construction (e.g., *She scared the hell out of me*) over successive time periods, by means of a multidimensional scaling algorithm (MDS; Kruskal 1964). MDS aims to place items in a 2-dimensional space such that the between-item distances are preserved as well as possible; in other words, it converts a set of distances between items into a set of coordinates for these items, which when plotted can be used to visualize distance relations. This method further offers the benefit of being entirely automatic and data-driven: no manual semantic intervention is required. In this paper, a similar method is used, but instead of MDS, it draws on a similar, more recent technique that achieves the same purpose: *t*-distributed Stochastic Neighbor Embedding (*t*-SNE for short; Van den Maaten & Hinton 2008). *T*-SNE avoids some pitfalls of MDS and was shown to be superior to it for dense spaces with many dimensions. As such it is particularly well suited to distributional data.

In both aim and method, this approach is similar to an earlier proposal by Gries & Stefanowitsch (2010), consisting in using hierarchical cluster analysis to identify semantic classes of verbs in the distribution of constructions on the basis of collocate frequencies. They also applied this method to the *way*-construction, and although their data come from a much smaller sample (namely the one-million-word ICE-GB), some of the classes they identify are similar to the ones found in the distributional semantic plots of the present study. However, one important difference between the two studies is that Gries & Stefanowitsch's method consists in clustering verbs according to the range of prepositions they occur with in the *way*-construction, while the present study

[8] It should be acknowledged that the difference with some of its competitors was small. In particular, the same model with all words instead of only content words rank as close second. This is an interesting and unexpected finding, which questions the importance of restricting the collocate search to content words, as is common practice in distributional semantics. It is fairly possible that dimensionality reduction and especially weighting make up for the noise that some function words might introduce.

uses collocation information collected in the entire corpus. Gries & Stefanowitsch's approach had to be limited to a selection of the most frequent items (32 verbs out of 492 in the case of the *way*-construction), which automatically rules out hapax legomena. The present method, on the other hand, allows to appreciate the distribution in its entirety, and to identify finer-grained verb classes.

In the following three sections, this method will be applied to the study of the *way*-construction, in particular with regard to the examination of the semantics of the verbs occurring in it at different points in time over the last 180 years. The three senses of the construction, path-creation, manner, and incidental-action, as described in Section 2, will be analyzed separately. The rationale for this decision is that the different senses of the construction also correspond to different roles played by the verb in the motion event described by the construction. Thus, the verbal distributions of the different senses are not strictly comparable. Besides, the different senses might well follow different paths of development, which would be obscured in a general analysis considering the construction as a single semantic unit.

The data for each sense of the *way*-construction were split into four time periods of roughly the same length (two periods of fifty years and two of forty years): 1830-1879, 1880-1929, 1930-1969 and 1970-2009. These divisions will be referred to as periods 1 to 4 in the following sections.

4.2. The path-creation sense

The distribution of the construction in each period is plotted in Figure 2. The location of each verb was determined by a *t*-SNE algorithm on the basis of the distributional model, with 5,000 iterations and a perplexity parameter of 10, using the "tsne" package in the R environment (R Development Core Team 2015). The perplexity parameter controls the extent to which the algorithm preserves low-scope similarity relations between items; the especially low value (typical values range between 5 and 50, cf. Van den Maaten & Hinton 2008) was chosen in order to emphasize tight clusters of highly similar items and identify semantic groups of verbs more easily. The size of the bullets in Figure 2 (and all other distributional semantic plots in this paper) gives an indication of the token frequency of each verb in the construction. The diameter of each bullet is proportional to the natural logarithm of the frequency; hence, size zero corresponds to hapax legomena. The color coding is not meant to be meaningful: it is merely intended as a visual aid to discern groups of items plotted close together. These groupings were determined automatically by a hierarchical clustering algorithm (using Ward's method) on the basis of the distance data returned by *t*-SNE, and many of them do correspond (or at least partly) to coherent semantic categories. In order to fit the page, the plots had to be scaled down and a particularly small font had to be used to avoid that the labels overlap too much. While the general tendencies should still be largely observable, some parts of the plots are unfortunately a little hard to read. High-resolution versions of all distributional semantic plots shown in this paper are provided in the online supplement, which can be consulted in particular to have a closer look at especially dense areas of the plots.

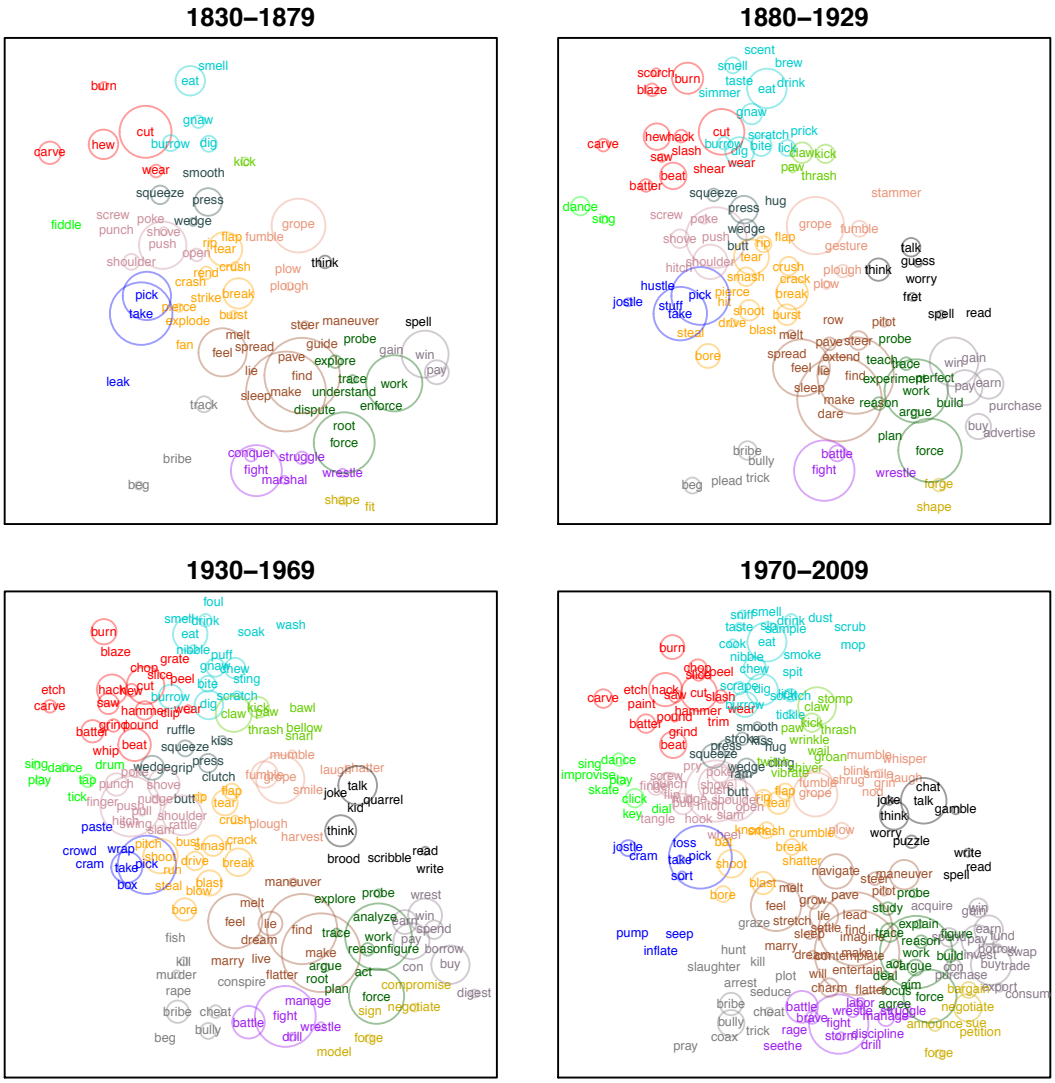

Figure 2: Distributional semantic plots of the path-creation sense of the *way*-construction.

First, the distributional semantic space reflects a general concrete/abstract distinction, with more concrete actions involving the manipulation of physical objects (e.g., *cut*, *dig*, *kiss*, *pierce*, *push*) found in the top left half, and more abstract ones found in the bottom right half (e.g., *bribe*, *marry*, *purchase*, *threaten*, *worry*). Interestingly, the general verbs *find* and *make* occur close together. This is probably due to the fact that these verbs are highly polysemous and semantically abstract, but it also lines up with the notion that they both do not contribute much to the sentence beyond the motion event conveyed by the construction. Among the concrete verbs, finer-grained groupings can be discerned. First, the middle area of the concrete verbs is occupied by groups of change-of-state verbs describing energetic actions, often resulting in negative effects on another entity: destruction (*blast*, *break*, *crush*, *smash*), harm (*beat*, *shoot*), separation (*cut*, *hack*, *hew*, *plow*, *saw*, *split*, *rip*, *tear*, *trim*), hitting (*claw*, *kick*, *paw*, *punch*, *slam*, *thrash*), pushing and pulling (*hitch*, *jostle*, *poke*, *pull*, *push*, *shove*, *shoulder*, *wedge*), or general damage (*batter*, *burn*, *hammer*, *pierce*, *strike*, *wear*). These groups, illustrated by examples (12a-f) below, are among the most populated at all times, and the semantics of their verbs is in line with the literal reading of the actual creation of a path that is argued both by Israel (1996) and Traugott & Trousdale (2013) to mark the origin of the path-creation sense of the *way*-

construction, with such precursor verbs as *clear*, *cut*, *dig*, *pave*, and *smooth*. This finding is reminiscent of Hopper's (1991: 22) principle of persistence in grammaticalization, according to which "when a form undergoes grammaticization from a lexical to a grammatical function, […] some traces of its original lexical meanings tend to adhere to it, and details of its lexical history may be reflected in constraints on its grammatical distribution".

(12) a. The steamers were using dynamite in an attempt to blast their way through the ice. (1931)

b. [T]hese men might […] try to shoot their way into Don Chafin's stronghold. (1978)

c. [N]ow cutting their way through the tangled woods, they toiled courageously forward. (1859)

d. He found that a bullet […] had punched its way through the cloth. (1867)

e. It occupied all Cap's faculties to push her way through the overhanging and interlacing branches of the trees. (1888)

f. He screamed in agony as the white metal burned its way through his flesh. (1961)

In general, the semantic domain of the construction essentially becomes more dense in areas that were already populated from the start. However, when the plot is examined more closely, it appears that this happens with some systematicity for particular semantic groups of verbs that are, however, not at all similar to the semantic core of forceful actions. At the very top of the plot, a loose group of verbs related to bodily functions regularly attracts new members from period 2 onwards, as exemplified by (13) below: ingestion (*eat*, *drink*, *nibble*, *puff*, *sample*, *sip*, *smoke*, *taste*), olfaction (*scent*, *smell*), and actions involving the mouth (*bite*, *chew*, *gnaw*, *spit*). Also arising in the same period towards the bottom right corner of the plot is a group of verbs having to do with commerce and finance (found in the bottom left corner of the plot and exemplified by [14] below) is growing from the end of the 19th century onwards: *acquire*, *advertise*, *borrow*, *buy*, *earn*, *export*, *fund*, *invest*, *pay*, *purchase*, *spend*, *squander*, and *swap*. Various other classes only become prominent towards the middle of the 20th century (period 3), in particular a set of overlapping clusters found between the two groups just mentioned, exemplified by (15) below: sound emission and manner of speaking (*bellow*, *groan*, *mumble*, *snarl*, *stammer*, *wail*, *whisper*), various kinds of social interaction (*chat*, *chatter*, *grin*, *joke*, *kid*, *laugh*, *nod*, *quarrel*, *smile*, *talk*), and a few cognitive verbs (*brood*, *puzzle*, *think*, *worry*).

(13) a. I waited until she had puffed her way to the end of the cigarette. (1942)

b. [H]e had almost chewed his way out of his wooden enclosure. (1958)

(14) a. I remember once telling a waiter how, with us, poor lads often earn their way through college by waiting on table at summer resorts. (1896)

b. There was an assumption you could sell stock to the public and fund your way through anything. (2003)

(15) a. Macmillan […] mumbled his way through a string of "Splendids," "Jolly goods," and "God bless you alls." (1958)

b. Duhamel grinned his way into the spotlight, chopping wood sans shirt and outshining his romantic foil, Topher Grace. (2006)

c. I was trying to think my way around the gun in his hand. (1949)

Another group arising more slowly around the same time towards the bottom left of the plot contains verbs describing various kinds of attack, coercion, and misconduct, as exemplified by (16): *bribe*, *bully*, *cheat*, *coax*, *conspire*, *hunt*, *kill*, *murder*, *plot*, *rape*, *seduce*, *slaughter*, and *trick*. Finally, many individual abstract verbs that were not attested in the first period but are found in the latest period (if not earlier) can be mentioned: *announce*, *argue*, *charm*, *entertain*, *explain*, *flatter*, *reason*, *read*, *study*, and *write*, to name only but a few.

(16) a. Nobody likes to think that the new vice president of the company murdered his way into the position. (1953)

b. Charlotte tried not to think of Space Station Freedom when two Senators had bullied their way onto the space station. (1996)

In sum, while the path-creation sense of the *way*-construction is still largely centered (although slightly decreasingly) on verbs describing forceful actions leading to the creation of an actual path through some concrete obstacle, in line with its historical origins, over the last two centuries it has come to recruit more and more verbs that do not typically describe ways in which a path can be created. This includes various types of abstract actions, but also concrete actions for which the enablement of motion is necessarily more indirect than literal path creation. This trend starts between the late 19th century and the middle of the 20th century (depending on the semantic class), and intensifies in the late 20th century.

4.3. The manner sense

The diachronic variation of the semantic distribution of the manner interpretation of the *way*-construction is plotted in Figure 3.[9] As reported in Section 3, the diachronic increase in type frequency is less steep for the manner sense than for the path-creation sense, so we expect qualitative developments to be more circumscribed. Not surprisingly given the semantics of the relevant sense, we find that most of these verbs encode some kind of manner of motion, which the distributional model seems to sort into groups whose members share particular semantic properties.

For instance, the verbs *edge*, *thread*, *trail*, *weave*, and *wind* appear close together arguably because they share the notion that motion occurs in a gradual or indirect way, likely due to the presence of obstacles. Almost all members of this group occur in all time periods, and some of them are among the most frequent verbs in the distribution. Another similar group that recurs with some frequency throughout the recent history of the construction contains verbs that describe difficult motion: *plod*, *toil*, *tramp*, and *trudge*. These groups, exemplified in (17), line up with what seems to be the preferred semantics of the construction, i.e., difficult motion despite hindrance (Goldberg 1995).

(17) a. She edged her way through the narrow space between her car and the next and climbed in behind the wheel. (1947)

b. The horse was plodding its way through the snow-drifts and it was evident that the animal would soon become exhausted. (1907)

[9] Some readers might wonder why at least two of the verbs typically cited in examples of the construction, *wend* and *worm*, are absent from the semantic plots of Figure 3. This is because these verbs were not frequent enough to reach the 1,000 frequency threshold required for inclusion in the distributional semantic model.

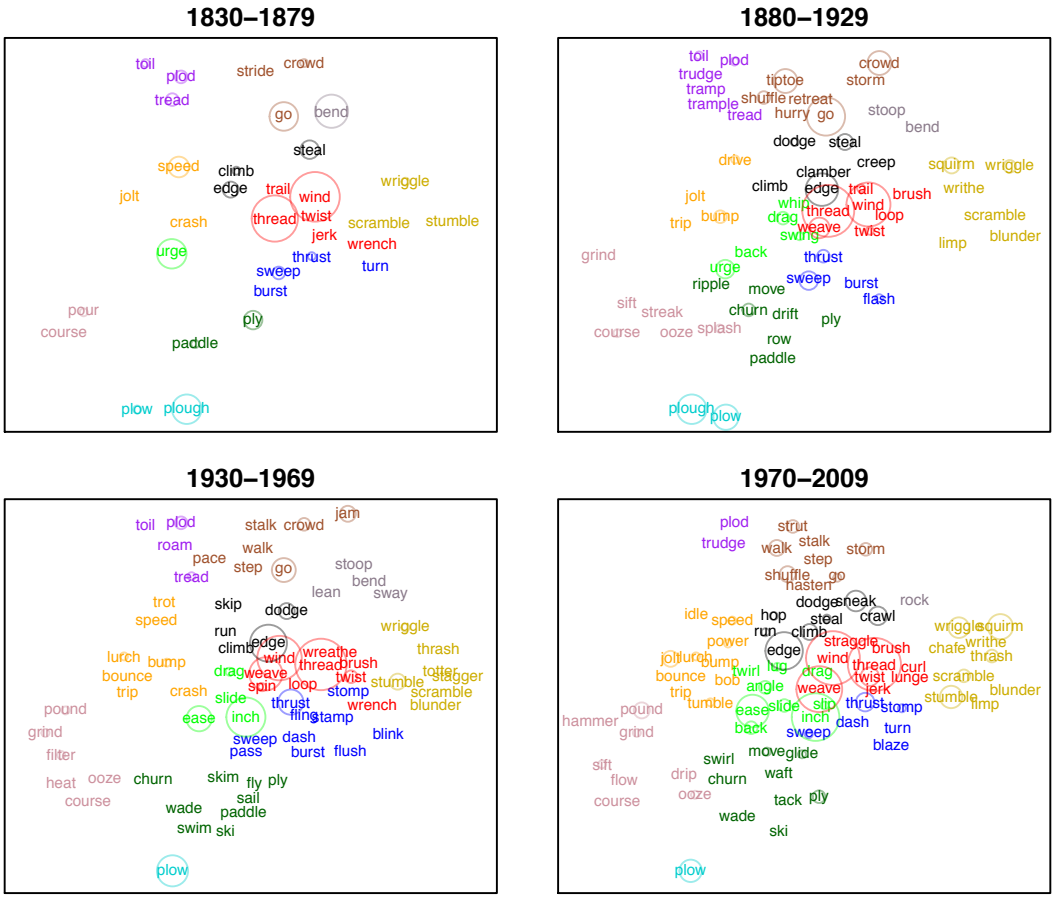

Figure 3: Distributional semantic plots of the manner sense of the *way*-construction.

Most of the other groups start with few members, but grow in size over time. One notable growth involves verbs encoding clumsy or unsteady motion (at the rightmost edge of the plot), usually due to the terrain or to the mover. *Scramble* and *stumble*, found at the outset, are later joined by other similar verbs, especially in the second half of the 20th century (periods 3 and 4): *blunder*, *bump*, *limp*, *totter*, *tumble*, *stagger*, and *trip*. This group is surrounded by verbs that primarily encode body movements: *bend*, *lean*, *squirm*, *stoop*, *thrash*, *twist*, *wrench*, *wriggle*, and *writhe*. In uses of these verbs in the construction, the body movement performed by the agent facilitates the execution of motion that was made difficult by some hindrance, as shown in sentence (18) below.

(18) Richie wriggled his way under the bed and groped along the floor until his hand touched the sharp corner of the box. (2004)

While the notion of difficult motion seems to be an important factor in the productivity of the manner *way*-construction, this does not prevent unrelated verbs from joining its distribution. From period 2 onwards, the construction gradually becomes more open to various types of motion verbs that do not involve difficulty, as exemplified by (19a-c) below: verbs of manner of walking in the upper left corner (*pace*, *stride*, *shuffle*, *tiptoe*, *trot*, *walk*, *step*, *stalk*, *strut*), verbs of rapid motion (*dash*, *hasten*, *power*, *run*, *speed*), verbs of aquatic motion (*swim*, *wade*), verbs of manner of motion specific to a certain kind of vehicle or theme (*drive*, *fly*, *glide*, *paddle*, *ply*, *row*, *sail*, *ski*, *tack*), and verbs of fluid motion (*course*, *drip*, *filter*, *ooze*, *pour*, *ripple*, *sift*, *splash*), all in the bottom half. None of these verbs necessarily involves difficult

motion *per se*, and some even encode rather neutral manners, if they do not actually indicate ease of motion.

(19) a. He slowly paced his way down the line of standing gangsters. (1951)

b. I must fly my way across the ocean. (1931)

c. Between them a pale shaft of sun managed to ooze its way through the grime on the glass panes. (1941)

To summarize, changes in the distribution of the manner sense of the *way*-construction are less significant than those of the path-creation sense. No strikingly new category of motion verbs seems to have joined the distribution of the construction, but some categories have grown more significantly than others. In particular, the construction is more likely to attract verbs that are compatible with the idea that motion is difficult in some way, although this tendency appears to be less pronounced in the more recent decades, which saw the occurrence of more neutral manners of motion in the construction.

4.4. The incidental-action sense

The semantic distribution of the incidental-action sense of the *way*-construction in diachrony is plotted in Figure 4.

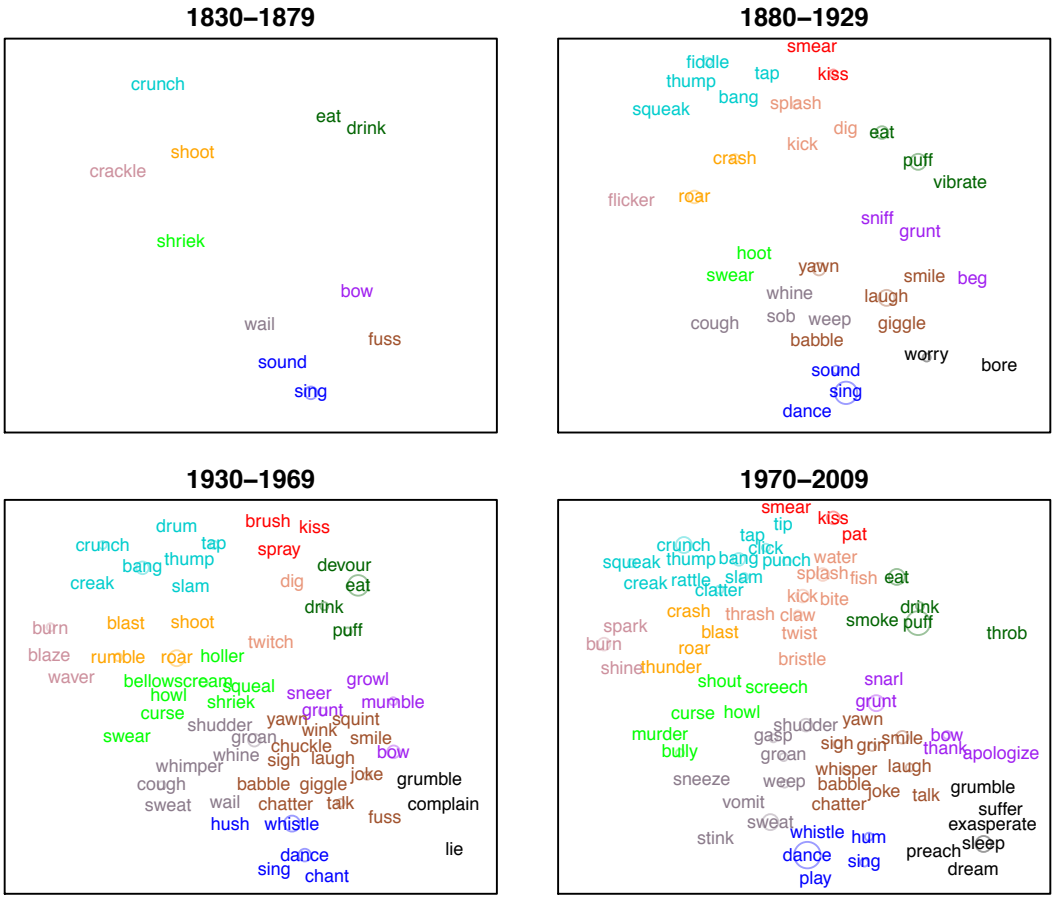

Figure 4: Distributional semantic plots of the incidental-action sense of the *way*-construction.

At the outset of the period of interest, the incidental-action sense is a recent innovation. Goldberg (1995: 203) and Israel (1996: 222) mention that the earliest example recorded in the OED (with the verb *ring*) dates back to 1836, and Traugott &

Trousdale (2013: 87) find that it is attested as early as 1820 (with the verb *shoot*). The earliest token I found in the entire COHA, reported in (20) below, is from 1839. This is the only example in the 1830s.

(20) This strangelooking personage began to bow his way through the crowd, with quick, nervous, hinge-like motions.

Unsurprisingly, this sense starts with few types in the first period, but they are already very diverse semantically. The semantic domain of the construction continues to expand in later periods, especially periods 3 and 4, as can be seen by the densification of the semantic plots in Figure 4. While many appear to be semantically isolated, quite a number of them can be grouped into several semantic classes, e.g., ingestion (*drink*, *eat*, *puff*, *smoke*, *sniff* and to some extent *smell*), performance (*chant*, *dance*, *hum*, *play*, *sing*, *sound*, *whistle*), bodily functions (*cough*, *sneeze*, *stink*, *sweat*, *vomit*), as exemplified by (21a-c). The distribution is clearly dominated by different types of verbs describing sound emission, be it noise (*bang*, *clatter*, *creak*, *crunch*, *drum*, *rattle*, *rumble*, *slam*, *squeak*, *tap*, *thump*), speech (*apologize*, *babble*, *chatter*, *curse*, *grumble*, *joke*, *swear*, *talk*, *thank*, *whisper*), cry (*bellow*, *holler*, *howl*, *roar*, *scream*, *screech*, *shout*, *shriek*, *squeal*), or other human sounds (*gasp*, *groan*, *grunt*, *hoot*, *sigh*, *snarl*, *sob*, *wail*, *weep*, *whimper*, *whine*, *yawn*), as illustrated by examples (22a-c).

(21) a. Timothy placidly ate his way through Paris. (1931)

b. [H]e spends about three months a year dancing his way around the globe on cruise ships. (2007)

c. Alfred Scanlon, with an air of gloom, deprecatingly coughed his way into the bedroom. (1921)

(22) a. [T]he tractor began to thump its way into the wheat belt. (1935)

b. Others died under his knife, or babbled their way to eternity in a fog of opium. (1995)

c. The children hooted and screeched their way from house to house. (1999)

The prominence of sound emission can be explained by the fact that it is quite amenable to being performed simultaneously with motion. Yet, the construction also occurs in early periods with other kinds of verbs that do clearly include any form of sound emission in their meaning, such as *bow*, *drink*, *eat*, and *fiddle*. Hence, the data do not fully bear out Israel's (1996: 225) claim that "[u]ntil well into the twentieth century instances of this sort consistently involve sounds produced in the process of moving". Unless the difference is due to some unexpected divergence with Israel's mostly British English data from the OED, this finding argues against Israel's proposal that the incidental-action sense might have emerged from an analogy between manner and the sound accompanying certain forms of motion, then extended to other kinds of verbs, since this explanation relies on the assumption that motion-related sound emission predated other meanings, which the COHA data shows to be unwarranted. This provides an argument in favor of Israel's second hypothesis, i.e., that the incidental-action use is an extension of the path-creation sense, drawing on the idea that it relates to cases where the enablement of motion is accidental rather than intended (Israel 1996: 225).

In the next section, I take stock of these findings and discuss them in relation to mechanisms of language change as they are envisaged in diachronic construction grammar.

4.5. Schematicity and productivity of the *way*-construction

As shown by the distributional semantic plots reported in the previous section, the distributions of verbs occurring in all three senses of the *way*-construction have undergone change over the past 180 years. This is worth noting given that the period of interest is both rather recent and not especially long with respect to the entire history of the construction. In this section, these patterns of change are interpreted in terms of the grammatical representation of each of the three senses in a diachronic construction grammar. In particular, the following discussion will focus on aspects of productivity and schematicity.

Productivity refers to knowledge about the range of lexical items that may be used in the slots of a construction. Schematicity refers to the level of detail with which a construction is represented, which in turn defines restrictions on the range of situations that its uses may apply to. Productivity and schematicity are commonly thought to be interrelated (Barðdal 2008; Hilpert 2012). Indeed, to the extent that more schematic constructions involve more schematic slots, they exert fewer constraints on the lexical items they can accommodate; hence, they tend, *ceteris paribus*, to be more productive. Conversely, insofar as a schematic slot captures the commonalities between items attested in it, the occurrence of a wider semantic range of items automatically makes the corresponding slot more schematic; novel items that are at odds with an existing schema might initially be perceived as deviant, but they can contribute to extending the schema if they gain acceptance.

The interdependency of schematicity and productivity can first be illustrated by the manner sense of the *way*-construction. As mentioned in Section 5.3, this sense appears to display a preference for verb meanings that emphasize the difficulty of motion. However, not only are verbs encoding more neutral manners of motion attested from the start, albeit scarcely, but such verbs also become more common in later decades, forming exemplar pockets of their own. Hence, the verb slot of the manner schema, which can plausibly be characterized as referring to difficult motion at the outset of the period, is becoming more abstract as more diverse verbs are used in the construction, and is changing towards a general manner-of-motion verb slot.

By contrast, the wide semantic diversity and the absence of sizeable verb clusters in the initial distribution of the incidental-action sense indicate that the verb slot most likely involves an abstract schema from the start, or at least very early on. Since the range of possible activities that can be performed during motion is virtually open-ended, while the range of actions that can be conducive to the creation of a path is at any rate more limited, the verb slot can be claimed to be highly schematic, and correspondingly, the incidental-action sense occurs with a wider semantic range of verbs than the other senses, although it is much more recent.

As indicated earlier, the recent productivity of the path-creation sense largely falls into groups sharing particular semantic properties, e.g., ingestion, commerce, misconduct, etc. Members of many of these groups were attested from the start (if only scarcely), which could mean that an increase in schematicity is not necessary to account for changes in the distribution: new members could be accounted for by being subsumed under low-level schemas, or by analogy with existing exemplars. On the other hand, the construction also attracts verbs that do not seem to be related to any other in the distribution, e.g., *grow*, *live*, *plan*. Besides, at the beginning of the period of interest, the distribution of the path-creation sense is already quite diverse semantically, which supports the suggestion that a schematic verb slot for the path-

creation sense was already in existence, licensing the occurrence of virtually any verb that can be construed as a means of path creation in a certain context. The existence of such a schema does not necessarily entail that all compatible items should be attested: like any product of linguistic usage, the observed productivity of a construction is subject to extra-linguistic factors relating to the communicative needs of speakers in particular contexts. A compelling example of this fact is found in the first three attestations of *spend* in the *way*-construction, all from the 1930s:

(23) a. Is it true that we can spend our way to prosperity? (1935)

b. There is no recorded instance of any nation having spent its way out of a depression. (1935)

c. [S]uch a statement stands in clear opposition to the Administration's philosophy of spending our way into recovery. (1939)

These sentences (all from the press, either newspapers or magazines) make reference to the welfare state policies enacted by Franklin D. Roosevelt in the United States in reaction to the Great Depression (aka the "New Deal"). The use of *spend* in the *way*-construction aptly captured the then fairly novel idea (cf. [23b]) that injecting public funds into the economy (for instance through public works projects) should help the country recover from its predicament. The same description will be reprised in later decades for similar situations, although not exclusively for those. This suggests that while the use of *spend* in the *way*-construction could already have been licensed by the constructional schema much earlier, it is the socio-historical context of the New Deal that effectively coined it.

While there is no clear indication that the verb slot itself has increased in schematicity, such a change seems very plausible for the motion component of the path-creation schema. As mentioned earlier and as already illustrated by some of the examples cited in Section 4.2, the motion of the subject referent in the present-day *way*-construction can correspond not only to actual motion in physical space, but also abstract events metaphorically construed as motion: for instance, becoming a member of an institution in (24a) below, delivering speech in (24b), mentally visualizing things in (24c), reading in (24d), and the passing of time in (24e).

(24) a. [T]hey talk about Uncle Paul having bought his way into the Senate!

b. He watched as his $5,000-a-day attorney fumbled his way through a closing argument.

c. [S]he fought against one particular ultra-sexy male image that tried forcing its way into her mind.

d. By the time he was four he could spell his way through his book with only occasional pauses for breath.

e. With his vision becoming rapidly more impaired, he made his way through three years at Texas Tech on innate ability and luck.

These few examples only present a small fraction of the range of meanings that can be encoded by the construction in the COHA, some of which fall under conventional metaphorical mappings (à la Lakoff & Johnson 1980) also instantiated by other expressions, while others seem to be one-shot creative expressions drawing on a novel metaphorical conceptualization. Crucially, some verb meanings are more likely to enable abstract motion than others. This is especially true of unusual means of physical path creation (e.g., *drink*, *sleep*, *stroke*, *tickle*) and in particular abstract

actions (e.g., *argue*, *bribe*, *negotiate*, *spend*), which characterize the recent productivity of the path-creation sense. Yet, it is not entirely clear whether these verbs were recruited into the distribution in order to encode abstract motion events, since both kinds are possible and attested in the corpus, as exemplified with *bribe* below, used with concrete motion in (25a) and abstract motion in (25b).

(25) a. Reporters used them to bribe their way into off-limits areas.

b. [H]e was carefully saving his money to bribe his way up to that exalted position.

To examine the relation between the productivity of the verb slot and the abstractness of the motion component, a different version of the distributional semantic plots was created that displays both kinds of information simultaneously. The entire dataset was annotated for the abstractness of the motion component: if a sentence conveyed the fact that the subject referent physically moves from one point to another in concrete space, motion was coded as concrete; it was coded as abstract otherwise, covering all the cases exemplified in (24a-e) above and many more. According to this definition, fictive motion (e.g., *The trail cleverly worked its way through the blackwater swamps*) was classified as abstract, since the scene, although physical, is static, and motion does not actually occur. As with the sense annotations, the reliability of the motion abstractness annotations was tested by asking the same annotator to provide this information for the 149 tokens that both they and the author annotated as instances of the path-creation sense. There was again high agreement (94.6%), confirmed by Cohen's Kappa ($\kappa = 0.863$, $z = 10.5$, $p = 0$).

The ratio of concrete vs. abstract uses for each verb in each of the four time periods was calculated and used to derive a color-coding scheme. Each point in the original distributional semantic plots was given a hue comprised between red and blue that was proportionally related to the concrete vs. abstract ratio, with bright red signaling a verb that is always used with concrete motion, deep blue signaling a verb that is always used with abstract motion, and shades of purple signaling verbs sharing their distribution between concrete and abstract motion to varying extents. The resulting plots are reported in Figure 5.

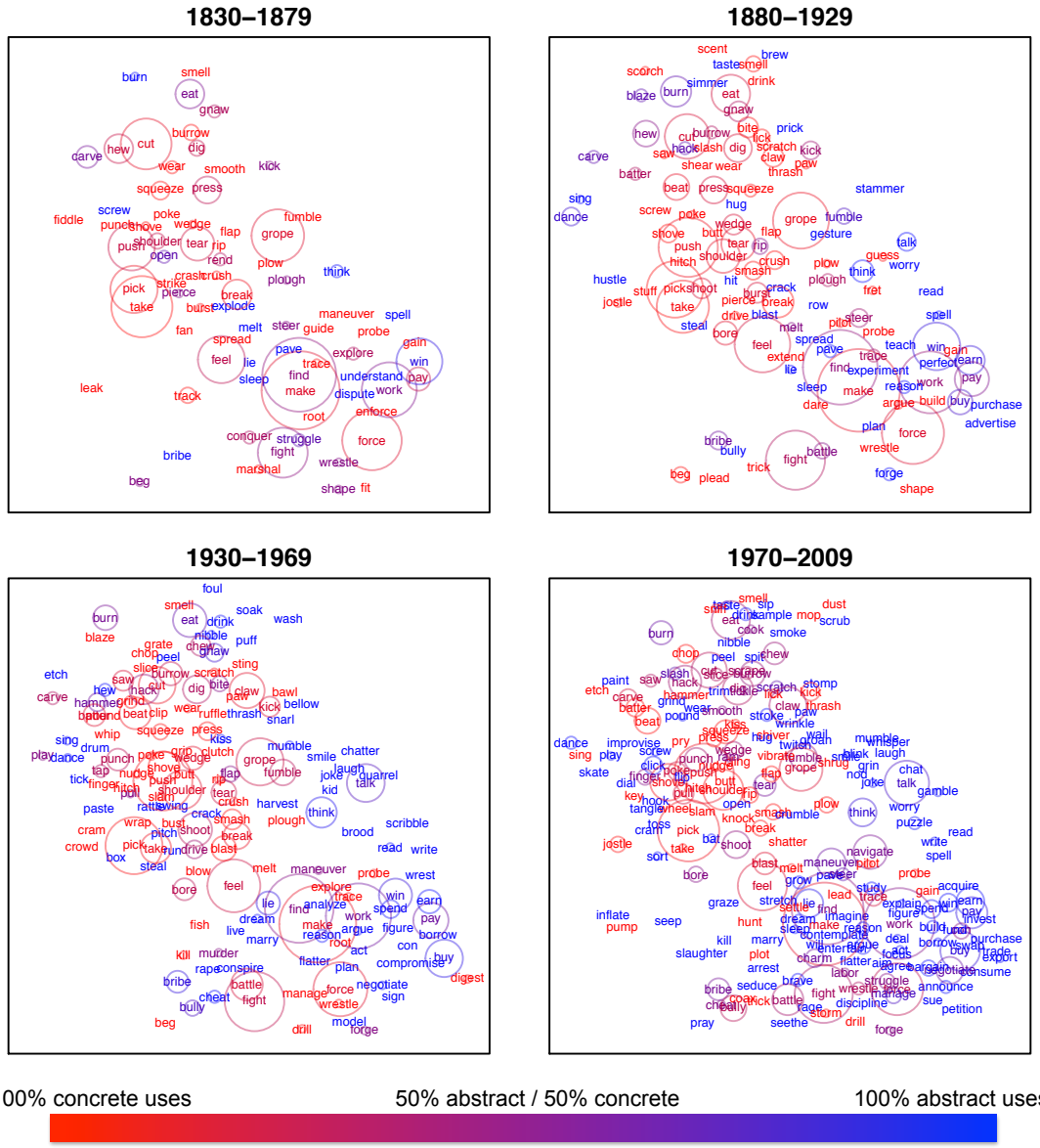

Figure 5: Distributional semantic plots of the path-creation sense of the *way*-construction, with color-coding to indicate the ratio of abstract vs. concrete motion uses for each verb.

At the beginning of the period of interest, most verb types are used exclusively or mostly with the expression of concrete motion; abstract motion is expressed by semantically general verbs (*find*, *work*) and only a handful of other types. Most of the "red" types correspond to what I characterized earlier as the semantic core of the construction: forceful, energetic actions prone to encode the literal creation of a physical path. Most of the blue types are located outside these core clusters, and quite spread out in the semantic space. In terms of grammatical representation, the structure of the semantic plot can be interpreted as reflecting a schema that is mostly centered on concrete motion; abstract types of motion are plausibly derived as metaphorical extensions of attested concrete uses with specific verbs (e.g., *fight*, *find*, *make*, *push*, *struggle*), but only sporadically with new types; abstract motion is not yet covered by the schema, or only marginally so. In later periods, the new semantic domains that are conquered by the construction, as reported in Section 4.2, largely consist of abstract uses. For instance, most of the new verbs in the social interaction group (*chat*, *joke*, *talk*) and in the commerce and finance group (*buy*, *pay*, *spend*) are "blue" (or "purple-

blue") types. This confirms that the semantic expansion of the construction into these regions of the semantic space is likely to be related to an increase in schematicity of the motion event it encodes, with the construction becoming more open to the expression of abstract motion. In future research, it could be interesting to have a closer look at the different types of abstract events that the construction encodes (a selection of which is presented in examples [24a-e]), and see if the range of possible abstract meanings also expands over time. This would provide further evidence for increased schematicity of motion in the *way*-construction.

5. Conclusion

This paper reported on a diachronic corpus-based study of the *way*-construction between 1830 and 2009, drawing on the COHA (Davies 2012). While the token frequency of the *way*-construction was found to be relatively stable over the period of interest, the number of different verbs attested in it has increased substantially over time. The paper focused on analyzing the productivity of the *way*-construction in semantic terms, addressing the question of what kinds of verb meanings joined the distribution, and when. To achieve this, a semantic characterization of the verbs was arrived at by means of a distributional semantic model. Drawing on the intuition that words with similar meanings tend to occur in similar contexts, a distributional semantic model captures the semantic similarity between words by comparing the frequency distribution of their collocates in a large corpus. The measure of semantic similarity thus derived was used to create distributional semantic plots of the *way*-construction: representations of how the semantic domain of the construction is populated at different points in time. Such plots were created for each of the three senses of the *way*-construction separately: the path-creation sense, the manner sense, and the incidental-action sense.

The examination of the semantic plots of the *way*-construction provided insights into recent change in the grammatical representation of the construction. It was found that significant developments took place in the distribution of each of the three senses of the construction. These changes were interpreted in terms of variation in the productivity and/or schematicity of the construction over time, in a diachronic construction grammar analysis (Hilpert 2013; Traugott & Trousdale 2013). For the manner sense, this was argued to be revealing of an increase in the schematicity of the verb slot, changing from verbs describing difficult motion to general manner of motion verbs. No such increase was found to be in evidence for the incidental-action sense, which was likely to be highly schematic from the start and has thus merely undergone a natural increase in productivity, with a particular prominence of verbs referring to sound emission. Finally, the path-creation sense was found to spread to increasingly more abstract kinds of verbs, which was argued to reflect an increase in the schematicity of the motion component, whereby the construction becomes more open to encoding abstract, metaphorical motion.

At the methodological level, one important contribution of this study was to showcase the potential of the distributional semantic approach for the study of grammatical constructions in diachrony, and their productivity in particular. The semantic classification offered by a state-of-the-art distributional semantic model is directly inferred from language use, and it is adequately reliable. This presents the advantage that no manual intervention is required to deal with the semantic aspects of the data,

which makes the method particularly efficient when dealing with large datasets containing dozens of lexical items, such as the one used for this paper. The case study of the *way*-construction illustrates how this method can be useful in informing hypotheses about changes in the schematicity of constructions by examining the entire set of items occurring in them and considering the identified trends of semantic change in light of the constructional meaning. This makes this method a valuable addition to the range of techniques available to scholars working in the framework of diachronic construction grammar from a quantitative and corpus-based perspective.

References

Andrews, Mark, Gabriella Vigliocco & David P. Vinson. 2009. Integrating Experiential and Distributional Data to Learn Semantic Representations. *Psychological Review* 116(3). 463–498.

Baayen, Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics. An International Handbook, Vol. 2*, 899–919. Berlin: Mouton De Gruyter.

Baayen, Harald & Rochelle Lieber. 1991. Productivity and English derivation: A corpus-based study. *Linguistics* 29. 801–844.

Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: John Benjamins.

Bergs, Alex & Gabriele Diewald (eds.). 2008. *Constructions and Language Change*. Berlin: Mouton de Gruyter.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.

Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, Joan & Clay Beckner. 2009. Usage-based Theory. In Heiko Narrog & Bernd Heine (eds.), *The Oxford Handbook of Linguistic Analysis*, 827–955. Oxford: Oxford University Press.

Bybee, Joan & David Eddington. 2006. A usage-based approach to Spanish verbs of 'becoming'. *Language* 82(2). 323–355.

Colleman, Timothy. 2011. Ditransitive Verbs and the Ditransitive Construction: A Diachronic Perspective. *Zeitschrift für Anglistik und Amerikanistik* 59(4). 387–410.

Colleman, Timothy & Bernard De Clerck. 2011. Constructional semantics on the move: On semantic specialization in the English double object construction. *Cognitive Linguistics* 22(1). 183–209.

Davies, Mark. 2012. Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. *Corpora* 7(2). 121–157.

Ellis, Nick C., Matthew O'Donnell & Ute Römer. 2014. Does language Zipf right along? Investigating robustness in the latent structures of usage and acquisition. In Jeffrey Connor-Linton & Luke W. Amoroso (eds.), *Measured language: quantitative studies of acquisition, assessment, processing and variation*, 33–50. Washington DC: Georgetown University Press.

Erk, Katrin. 2012. Vector Space Models of Word Meaning and Phrase Meaning: A Survey. *Language and Linguistics Compass* 6(10). 635–653.

Evert, Stefan. 2005. *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. Stuttgart: Institut für maschinelle Sprachverarbeitung, University of Stuttgart. urn:nbn:de:bsz:93-opus-23714.

Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*. *Language* 64(3). 501–538

Firth, John R. 1957. A synopsis of linguistic theory 1930-1955. In *Studies in linguistic analysis (Special volume of the Philological Society)*, 1–32. Oxford: Blackwell.

Fried, Mirjam. 2009. Construction Grammar as a Tool for Diachronic Analysis. *Constructions and Frames* 1(2). 261–91.

Fried, Mirjam. 2010. Grammaticalization and Lexicalization Effects in Participial Morphology: A Construction Grammar Approach to Language Change. In An Van linden, Jean-Christophe Verstraete & Kristin Davidse (eds.), *Formal Evidence in Grammaticalization Research*, 191–224. Amsterdam: John Benjamins.

Fried, Mirjam. 2013. Principles of constructional change. In Thomas Hoffmann and Graeme Trousdale (eds.), *The Oxford handbook of Construction Grammar*, 419–437. Oxford: Oxford University Press.

Fried, Mirjam & Jan-Ola Östman. 2004. Construction grammar: A thumbnail sketch. In Mirjam Fried & Jan-Ola Östman (eds.), *Construction grammar in a cross-language perspective*, 11–86. Amsterdam: John Benjamins.

Gaeta, Livio & Davide Ricca. 2006. Productivity in Italian word formation: A variable-corpus approach. *Linguistics* 44(1). 57–89.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, Adele E., Devin M. Casenhiser & Nitya Sethuraman. 2004. Learning argument structure generalizations. *Cognitive Linguistics* 15(3). 289–316.

Gries, Stefan T. & Anatol Stefanowitsch. 2010. Cluster analysis and the identification of collexeme classes. In Sally Rice & John Newman (eds.), *Empirical and experimental methods in cognitive/functional research*, 73–90. Stanford, CA: CSLI.

Hill, Felix, Roi Reichart & Anna Korhonen. 2014. SimLex-999: Evaluating Semantic Models with (Genuine) Similarity Estimation. URN: arXiv:1408.3456 <http://arxiv.org/abs/1408.3456v1> (15 September, 2015).

Hilpert, Martin. 2012. Diachronic collostructional analysis. How to use it, and how to deal with confounding factors. In Kathryn Allan & Justyna Robynson (eds.), *Current Methods in Historical Semantics*, 133-160. Berlin: Mouton de Gruyter.

Hilpert, Martin. 2013. *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax*. Cambridge: Cambridge University Press.

Hopper, Paul J. 1991. On some principles of grammaticalization. In Elizabeth C. Traugott & Bernd Heine (eds.), *Approaches to Grammaticalization, Vol. 1*, 17–35. Amsterdam: John Benjamins.

Israel, Michael. 1996. The way constructions grow. In Adele E. Goldberg (ed.), *Conceptual structure, discourse and language*, 217–230. Stanford, CA: CSLI Publications.

Jackendoff, Ray. 1990. *Semantic structures*. Cambridge, MA: MIT Press.

Kay, Paul & Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: The What's X doing Y? construction. *Language* 75. 1–33.

Kruskal, Joseph B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29(1). 1–27.

Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

Landauer, Thomas K., Peter W. Foltz & Darrell Laham. 1998. Introduction to Latent Semantic Analysis. *Discourse Processes* 25. 259–284.

Lenci, Alessandro. 2008. Distributional semantics in linguistic and cognitive research. *Rivista di Linguistica* 20(1). 1–31.

Lund, Kevin, Curt Burgess & Ruth A. Atchley. 1995. Semantic and associative priming in a high-dimensional semantic space. In *Cognitive Science Proceedings (LEA)*, 660–665.

Mondorf, Britta. 2011. Variation and change in English resultative constructions. *Language Variation and Change* 22(3). 397–421.

Noël, Dirk. 2007. Diachronic construction grammar and grammaticalization theory. *Functions of Language* 14(2). 177–202.

Noël, Dirk. 2008. The Nominative and Infinitive in Late Modern English: A Diachronic Constructionist Approach. *Journal of English Linguistics* 36(4). 314–340.

Noël, Dirk & Timothy Colleman. 2010. *Believe*-type raising-to-object and raising-to-subject verbs in English and Dutch: A contrastive investigation in diachronic construction grammar. *International Journal of Corpus Linguistics* 15(2). 157–182.

Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54(1). 149–188.

Perek, Florent & Maarten Lemmens. 2010. Getting at the meaning of the English *at*-construction: the case of a constructional split. *CogniTextes* 5. URL: http://cognitextes.revues.org/331

R Development Core Team. 2015. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. URL: http://www.R-project.org/

Römer, Ute, Matthew B. O'Donnell & Nick C. Ellis. 2015. Using COBUILD grammar patterns for a large-scale analysis of verb-argument constructions: Exploring corpus data and speaker knowledge. In Nicholas Groom, Maggie Charles & Suganthi John (eds.), *Corpora, Grammar and Discourse: In Honour of Susan Hunston*, 43–72. Amsterdam: John Benjamins.

Suttle, Laura & Adele E. Goldberg. 2011. The partial productivity of constructions as induction. *Linguistics* 49(6). 1237–1269.

Szczesniak, Konrad. 2013. You can't cry your way to candy: Motion events and paths in the x's way construction. *Cognitive Linguistics* 24(1). 159–194.

Traugott, Elizabeth C. 2003. Constructions in grammaticalization. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 624–647. Oxford: Blackwell.

Traugott, Elizabeth C. 2008. Grammaticalization, Constructions and the Incremental Development of Language: Suggestions from the Development of Degree Modifiers in English. In Regine Eckardt, Gerhard Jäger & Tonjes Veenstra (eds.), *Variation, Selection, Development: Probing the Evolutionary Model of Language Change*, 219–250. Berlin: Mouton de Gruyter.

Traugott, Elizabeth C. & Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.

Trousdale, Graeme 2010. Issues in Constructional Approaches to Grammaticalization in English. In Katharine Stathi, Elke Gehweiler & Ekkehard König (eds.), *Grammaticalization: Current Views and Issues*, 51–72. Amsterdam: Benjamins.

Trousdale, Graeme 2012. Grammaticalization, Constructions, and the Grammaticalization of Constructions. In Kristin Davidse, Tine Breban, Lieselotte Brems & Tanja Mortelmans (eds.), *Grammaticalization and Language Change: New Reflections*, 167–98. Amsterdam: Benjamins.

Turney, Peter & Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37. 141–188.

Van der Maaten, Laurens & Geoffrey Hinton. 2008. Visualizing Data using *t*-SNE. *Journal of Machine Learning Research* 9. 2579–2605.

Zipf, George K. 1935. *The psycho-biology of language: An introduction to dynamic philology*. Cambridge, MA: M.I.T. Press.